Spline Smoothing over Difficult regions

Koopman, S.J.; Wong, S.Y.

2008

document version
Early version, also known as pre-print

TI 2008-114/4
Tinbergen Institute Discussion Paper

Spline Smoothing over Difficult Regions

*Siem Jan Koopman**
Soon Yip Wong

Department of Econometrics, VU University Amsterdam.

** Tinbergen Institute.*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Spline Smoothing over Difficult Regions: a State Space Approach

Siem Jan Koopman[*] *and Soon Yip Wong*

Department of Econometrics, VU University Amsterdam

De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

Abstract

We consider the problem of smoothing data on two-dimensional grids with holes or gaps. Such grids are often referred to as difficult regions. Since the data is not observed on these locations, the gap is not part of the domain. We cannot apply standard smoothing methods since they smooth over and across difficult regions. More unfavorable properties of standard smoothers become visible when the data is observed on an irregular grid in a non-rectangular domain. In this paper, we adopt smoothing spline methods within a state space framework to smooth data on one- or two-dimensional grids with difficult regions. We make a distinction between two types of missing observations to handle the irregularity of the grid and to ensure that no smoothing takes place over and across the difficult region. For smoothing on two-dimensional grids, we introduce a two-step spline smoothing method. The proposed solution applies to all smoothing methods that can be represented in a state space framework. We illustrate our methods for three different cases of interest.

Keywords: Bivariate smoothing; Geo-statistics; Missing observations; Smoothing spline model; State space methods

JEL classification: C13, C22, C32

[*]Corresponding author: Siem Jan Koopman, Department of Econometrics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Email: s.j.koopman@feweb.vu.nl

1 Introduction

In recent years, there is a growing interest in designing methods to smooth data on one- or two-dimensional grids which contain holes or gaps. Surfaces with holes are referred to as *difficult regions* and are often associated with geographical locations. In this paper we propose new solutions for three problems in smoothing data over difficult regions. Firstly, obtaining smoothed estimates over a difficult region is not desirable, since the gap is not part of the domain and no information is available in this area. Standard smoothing methods however still produce estimates over the difficult region. As a result, most of these methods cannot be applied unless we modify them in an ad-hoc manner. Secondly, standard methods tend to smooth across the gap. In most cases it is not preferred to let smoothed values on one side of the gap depend on observations from the other side. Thirdly, when the grid is irregular or non-rectangular, most of the smoothing methods are inapplicable as they require the data to be equidistant and observed on a rectangular domain. Regular and rectangular domains are not common in practice.

Popular surface smoothing techniques are wavelet smoothing methods, kernel smoothing, spline smoothing and kriging. Wavelet smoothing is removing high frequencies from the data by decomposing it into a family of so-called analyzing signals, see Horgan (1999). Kernel smoothing is based on a kernel function that provide observation weights to obtain smoothed estimates of the data. We refer to Wand & Jones (1995) for an overview. Spline smoothing minimizes the squared distance between observed data and a spline function subject to a roughness penalty, see Kohn & Ansley (1987), Wahba (1990), Hastie & Tibshirani (1990) and Green & Silverman (1994). A method from the geo-statistics literature is kriging and is based on a linear least squares technique to estimate smooth functions on grids using information contained in observed data, see amongst others Cressie (1993).

When the grid is regular and it contains no difficult regions, the aforementioned methods can be applied straightforwardly. However, problems arise when the grid is irregular or non-rectangular or contains gaps in the domain. Wavelet smoothing cannot be applied when the grid is not regular, whereas spline smoothing, kernel smoothing and kriging share the problem that they cannot smooth around holes. We may employ methods such as finite element L-

splines, see e.g. Ramsay (2002) or low-rank thin plate splines, see e.g. Kammann & Wand (2003) and Wang & Ranalli (2007). The method of finite element L-splines partitions the grid and constructs a polynomial function on each piece of the partition such that the union of these pieces approximates the real function. Thin plate splines are the two-dimensional analogue of the one-dimensional cubic splines. The low-rank thin plate splines rely on the Euclidean distance between geographical locations as a measure of similarity among data points and construct surface approximations through it. This method has become more popular since it reduces the computation time to smooth the surface.

The key to our solutions of the three aforementioned problems is the straightforward treatment of missing observations by means of the state space framework. The first type of missing observations is used to solve the problem of the irregularity of the grid. We add pseudo points to the grid to create a regular domain. We consider a second type of missing observations to handle the presence of the difficult region, such that the spline model will not smooth across or over the difficult region. Moreover, in the case of two-dimensional grids, we transform non-rectangular grids into rectangular ones by adopting this second type of missing observations.

The introduction of the two types of missing observations leads to a mild reformulation of the cubic spline model in state space form. For smoothing on two-dimensional grids, we introduce a two-step smoothing method. The data is smoothed in one dimension first. In the second step, smoothing takes place in the other dimension using the smoothed data of the previous step. When a difficult region is present in the two-dimensional domain, we show that the same reformulations as we have in the univariate case, can solve the smoothing problems. The contribution of this paper is that issues related to smoothing over difficult regions are solved by introducing two types of missing observations and applying state space methods. By adding these missing observations and by reformulating the model, we obtain smoothed estimates of the data in a straightforward way.

The paper is organized as follows: section 2 deals with the smoothing methods on one- and two-dimensional grids. In section 3, we introduce two types of missing observations and we explain how our approach is able to solve the smoothing problems. In section 4, we present illustrations for both the univariate and bivariate cases. It is shown that our general methodology is effective in handling difficult regions for smoothing. Section 5 concludes.

2 Smoothing Methods

2.1 One-dimensional grids: cubic spline smoothing

Suppose we have a univariate sequential ordered series $y_1, \ldots, y_n$ for which its ith value y_i is observed at location τ_i for a one-dimensional domain τ. Values are observed at the grid $\tau_1, \ldots, \tau_n$. The observations are not necessarily equispaced and the distance between the observations are denoted by $\delta_i = \tau_{i+1} - \tau_i$. We wish to approximate the series by a smooth continuous function $\mu(\tau)$, i.e. $y_i = \mu(\tau_i) + \epsilon_i$, where ϵ_i is the measurement error. The common approach to smoothing is to choose $\mu(\tau)$ by minimizing

$$\sum_{i=1}^{n} [y_i - \mu(\tau_i)]^2 + \lambda \int \left[\frac{\partial^2 \mu(\tau)}{\partial \tau^2}\right]^2 d\tau, \qquad (1)$$

with respect to $\mu(\tau)$ and with a known value for the smoothing parameter $\lambda > 0$. The function $\mu(\tau)$ is referred to as a cubic spline function, see Wecker & Ansley (1983), Kohn & Ansley (1987), Wahba (1990), Hastie & Tibshirani (1990) and Green & Silverman (1994). The first term measures the fit to the data and the second term is a roughness penalty. The smoothing parameter λ controls the trade-off between fit and smoothness. Small values of λ produce spline estimates which fit the data better, while larger values result in smoother spline estimates. The connection between splines and the state space framework is discussed in Wecker & Ansley (1983) and Ansley et al. (1992). When we take the continuous process $\mu(\tau)$ at discrete intervals, we have the following model for the cubic spline

$$\begin{aligned} y_i &= \mu(\tau_i) + \epsilon_i, & \epsilon_i &\sim \text{NID}(0, \sigma_\epsilon^2), \qquad i = 1, \ldots, n, \\ \mu(\tau_{i+1}) &= \mu(\tau_i) + \delta_i \beta(\tau_i) + \xi_i, & \xi_i &\sim \text{NID}(0, \sigma_\zeta^2 \delta_i^3/3), \\ \beta(\tau_{i+1}) &= \beta(\tau_i) + \zeta_i, & \zeta_i &\sim \text{NID}(0, \sigma_\zeta^2 \delta_i), \end{aligned} \qquad (2)$$

where ϵ_i is uncorrelated with both ξ_i and ζ_i, but the error terms ξ_i and ζ_i are correlated with $E(\xi_i \zeta_i) = \sigma_\zeta^2 \delta_i^2/2$. The smoothing parameter λ in (1) is given by the ratio $\lambda = \sigma_\epsilon^2/\sigma_\zeta^2$.

A general representation of models such as the one in (2) is provided by the state space model, see Durbin & Koopman (2001) for a general treatment. The state space form of the model in (2) is given by:

$$y_i = Z_i \alpha_i + \epsilon_i, \qquad \epsilon_i \sim \mathrm{NID}(0, G_i), \qquad i = 1, \ldots, n, \tag{3}$$

$$\alpha_{i+1} = T_i \alpha_i + \eta_i, \qquad \eta_i \sim \mathrm{NID}(0, H_i), \tag{4}$$

where $\alpha_i = \{\mu(\tau_i), \beta(\tau_i)\}'$ is the state vector, ϵ_i is the measurement error and $\eta_i = \{\xi_i, \zeta_i\}'$ is the disturbance vector of the state equation (4). The measurement equation (3) relates the observations y_i to the state vector α_i through the signal $Z_i \alpha_i$ and the disturbance ϵ_i. The state vector α_i in (4) is a first order vector autoregressive process with a diffuse initialisation, that is $\alpha_1 \sim \mathrm{N}(0, \kappa I)$, $\kappa \to \infty$. The system matrices of the state space form are Z_i, transition matrix T_i and the variance matrices G_i and H_i which are given by

$$Z_i = \begin{pmatrix} 1 & 0 \end{pmatrix}, \qquad G_i = \sigma_\epsilon^2, \qquad T_i = \begin{pmatrix} 1 & \delta_i \\ 0 & 1 \end{pmatrix}, \qquad H_i = \begin{pmatrix} \sigma_\zeta^2 \delta_i^3 / 3 & \sigma_\zeta^2 \delta_i^2 / 2 \\ \sigma_\zeta^2 \delta_i^2 / 2 & \sigma_\zeta^2 \delta_i \end{pmatrix}.$$

The conditional mean and the variance of the spline $\mu(\tau)$ given the data are obtained by applying the Kalman filter and smoother. We refer to Appendix A for details. The conditional mean is the smoothed spline estimate. The parameters σ_ϵ^2 and σ_ζ^2 are estimated by maximizing the likelihood function, which is based on the Kalman recursions, see Schweppe (1965) and Harvey (1989, sec. 3.4). Since $\lambda = \sigma_\epsilon^2 / \sigma_\zeta^2$, the method of maximum likelihood provides a convenient alternative for the estimation of λ by cross-validation as discussed in Green & Silverman (1994). The relation between maximum likelihood and cross-validation estimation is explored in de Jong (1988) and Ansley et al. (1991).

For the computation of the filtered and smoothed state estimates, we implicitly assign weights to all observations of the sample. The Kalman recursions can be adapted to compute the weighting pattern, see Koopman & Harvey (2003). The filtered state $a_{i|i-1}$ and the smoothed state $a_{i|n}$ can be written as

$$a_{i|i-1} = \sum_{t=1}^{i-1} w_t(a_{i|i-1}) y_t, \qquad a_{i|n} = \sum_{t=1}^{n} w_t(a_{i|n}) y_t, \quad i = 1, \ldots, n, \tag{5}$$

where the weights $w_t(a_{i|i-1})$ and $w_t(a_{i|n})$ are assigned to observation y_t. Plots of the weighting pattern can be informative since it reveals how the spline estimates are constructed. It also enables the comparison with all kernel functions used in non-parametric smoothing methods, see Green & Silverman (1994).

2.2 Two-dimensional grids: two-step smoothing

To estimate smooth functions of data on two-dimensional grids, we use a simple two-step method based on the single spline function as discussed in the previous section. Let y denote the observed values of a unknown two-dimensional function $\mu(\tau, \omega)$. The relation between y and the function $\mu(\tau, \omega)$ is given by

$$y_{ij} = \mu(\tau_i, \omega_j) + \epsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m, \tag{6}$$

where y_{ij} is the observation for gridpoint (i, j) and ϵ_{ij} is the corresponding measurement error. In Figure 1, we depict a two-dimensional $n \times m$ grid, where the horizontal axis represents the τ-domain and the vertical axis corresponds to the ω-domain. The rows of the grid are denoted by $\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_m$, whereas the columns are denoted by $\boldsymbol{\tau}_1, \dots \boldsymbol{\tau}_n$. We propose a two-step method to estimate the unknown function $\mu(\cdot)$.

1. In the first step we smooth the data along the τ-domain. On each slice $\boldsymbol{\omega}_j$, for $j = 1, \dots, m$, we estimate single smooth spline functions as described in section 2.1. Smoothing takes place on the basis of model (2) with a known smoothing parameter λ which we denote by $\lambda_{\omega,j}$. The resulting m smoothed splines are stored as data to be used in the second step.

2. In the second step, we estimate spline functions along the w-domain. On each slice $\boldsymbol{\tau}_i$, for $i = 1, \dots, n$, we replace the data by the their smoothed estimates from the first step. The smoothing parameters are denoted by $\lambda_{\tau,i}$. The resulting n smoothed splines obtained in this step are our final smoothed estimate of the two-dimensional function $\mu(\cdot)$.

The basic two-step smoothing method for two-dimensional surfaces is computationally fast and easy to implement. The application of $m + n$ univariate smoothing operations is very fast nowadays, even on standard desktop computers. The decision to start with smoothing in the τ-domain in the first step may appear somewhat ad-hoc. However, whether we start with the τ-domain and then the ω-domain or vice-versa, the resulting spline estimates are similar. The two-step method is flexible, since for each slice in both the τ- and ω-domain, a different smoothing parameter can be chosen or estimated by the method of maximum likelihood. In

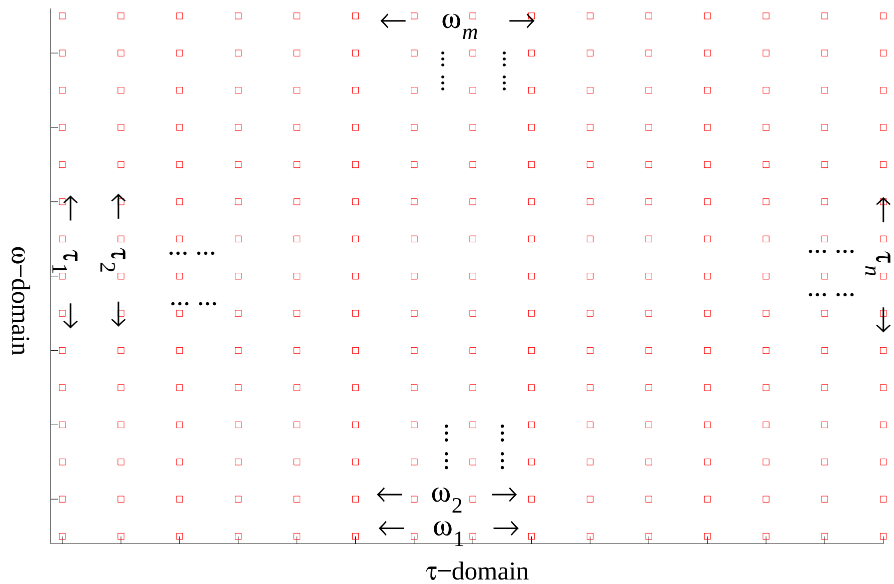

Figure 1: Two-dimensional grid. The horizontal axis represents the τ-domain and the vertical axis corresponds to the ω-domain.

our applications we have estimated the smoothing parameters for each smoothing operation by maximum likelihood. It has turned out that the estimates do not vary much locally. Over a wider set of slices, the smoothing parameter adapts slowly to different values. We regard this flexibility as a merit to our approach of two-dimensional smoothing.

The implied kernel function for our two-dimensional spline smoothing method is obtained as follows. By storing the weights in the two steps of our procedure, we obtain the weighting pattern for smoothing. Suppose we have a $n \times m$ grid and we are interested in the weighting pattern to obtain the smoothed value at (τ^*, ω^*). For this purpose, we have the following procedure:

1. We first compute the weighting patterns when we smooth along the τ-domain, i.e. the weights corresponding to smoothing at the points $(\tau^*, 1), (\tau^*, 2), \ldots, (\tau^*, m)$ are computed and stored. This means that for point (τ^*, j), we calculate the weights assigned to the observations at slice $\boldsymbol{\omega}_j$ for $j = 1, \ldots, m$.

2. Secondly, when we smooth along the ω-domain, we compute the weighting pattern at the points $(1, \omega^*), (2, \omega^*), \ldots, (n, \omega^*)$. For point (i, ω^*), we calculate the weights assigned to observations at slice $\boldsymbol{\tau}_i$ for $i = 1, \ldots, n$. By combining the two weighting

patterns, we obtain the weights for smoothing on two-dimensional grids.

In Figure 2, we present the two-dimensional weighting patterns for smoothing conditional on high and low values of the smoothing parameters. In panel (i) we present the case where $\lambda_{\tau,i} = \lambda_{\omega,j} = 50$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Since the smoothing parameters have relatively small values, the weights decay quickly once observations are away from the location of interest. The weighting pattern is concentrated around the point (τ^*, ω^*). In panel (ii), we have a relatively large value for the smoothing parameters on the slices $\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_m$, i.e. $\lambda_{\omega,j} = 1000$. The smoothing parameters on the other slices $\lambda_{\tau,i}$ are equal to 50. We see that more observations on the $\boldsymbol{\omega}_j$ slices are used to form the estimates, while non-zero weights are assigned to observations associated with the $\boldsymbol{\tau}_i$ slices which are relatively close to τ^*. Panel (iii) depicts the reverse case where $\lambda_{\omega,j} = 50$ and $\lambda_{\tau,i} = 1000$. Non-zero weights are now assigned to many observations on the slices $\boldsymbol{\tau}_i$, whereas a few observations on $\boldsymbol{\omega}_j$ are used. In panel (iv), all smoothing parameters are equal to 1000. Many observations around the point (τ^*, ω^*) are used to calculate smoothed estimates.

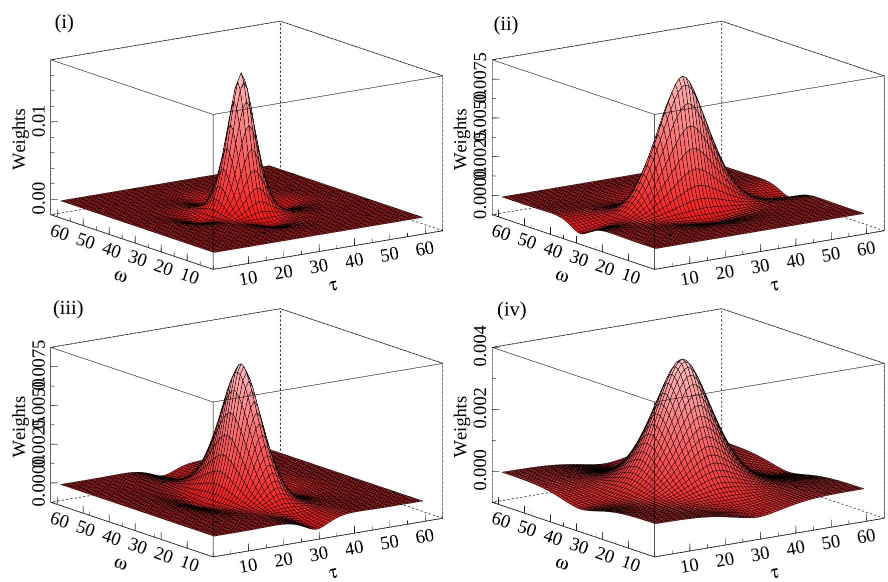

Figure 2: Weighting pattern for different values for smoothing parameters. Panel (i) depicts the case where $\lambda_{\tau,i} = \lambda_{\omega,j} = 50$, $i = 1, \ldots, n$, $j = 1, \ldots, m$. In panel (ii) we have $\lambda_{\tau,i} = 1000$ and $\lambda_{\omega,j} = 50$. In panel (iii) we have $\lambda_{\tau,i} = 50$ and $\lambda_{\omega,j} = 1000$. Panel (iv) depicts the case where $\lambda_{\tau,i} = \lambda_{\omega,j} = 1000$.

We present the two-step smoothing method as an alternative to the Smoothing Spline Analysis of Variance (SS–ANOVA) model, see Gu & Wahba (1993), Gu (2002) and Guo (2002). The SS–ANOVA model provides a framework for multivariate function estimation that allows for both main effects and interaction terms. It is widely used in biomedical studies to obtain smoothed estimates of data observed on two-dimensional grids, see amongst others Gao et al. (2001), Guo et al. (2003) and Wang et al. (2003). Qin & Guo (2006) show that this parametric smoothing spline model can be cast into state space form. Model parameters are estimated via the method of maximum likelihood. However, in many cases the weighting pattern for computing smoothed estimates does not assign nonzero values to data points located close to the point of interest, i.e. we only use observations in the τ-domain to calculate smoothed values. Many observations in the ω-domain are not used for this purpose. The two-step smoothing spline method proposed in this paper does not have this weighting problem since it assigns nonzero weights to many observations in both the τ-domain and the ω-domain.

3 Irregular Grids and Difficult Regions

3.1 Irregular grids

In applied work it is common to deal with data that is not observed at all points in the grid. When no values are recorded at points which are not part of the domain, we have a difficult region in the domain. The case where the data is not observed on an equispaced grid, is referred to as an irregular grid. It is not straightforward to smooth data when the grid is irregular or contains difficult regions, see the discussion by Ramsay (2002). The standard methods of smoothing require the data to be equispaced and free of gaps in the domain. We tackle the problem of unequally spaced data by introducing *missing observations of type 1*. Pseudo points are added to the grid such that the new series is equispaced. Denote D_1 as the collection of the indices which correspond to the missing observations of type 1. We emphasize that estimates should be obtained for these points.

For the univariate cubic spline model, the interpolation or extrapolation over the pseudo points in D_1 is equivalent to setting G_t, the variance of the measurement error ϵ_j in (3), equal to κ, where $\kappa \to \infty$, that is

$$G_t = \begin{cases} \kappa, & \text{for } t \in D_1, \\ \sigma_\epsilon^2 & \text{for } t \notin D_1. \end{cases} \tag{7}$$

The idea behind this modification is that although there is no information available for the measurement equation, the state vector still needs to be updated. As a result, we obtain estimates of the spline function for D_1. Moreover, we should assign zero-valued weights to the points corresponding to the missing observations of type 1 to form the filtered and smoothed state at index i, for $i = 1, \ldots, n$, since there is no information available coming from these points in D_1, i.e.

$$w_t(a_{i|i-1}) = 0, \qquad w_t(a_{i|n}) = 0, \qquad t \in D_1, \quad i = 1, \ldots, n. \tag{8}$$

We show in Appendix B that the modification in (7) leads to weighting patterns with zero weights for the points corresponding to missing observations of type 1.

For smoothing on two-dimensional grids, we require the same modifications as for the univariate spline model. By adding missing observations of type 1, we can create a regular grid. In both steps of the method described in subsection 2.2, we assign zero-valued weights to missing observations of type 1. The convolution of the weighting patterns resulting from the two steps (in which the treatment of missing observations is exactly the same) ensures that in the final weighting pattern, zero-valued weights are assigned to all missing observations of type 1. This implies that estimates can be obtained for such missing points, while they are not used to calculate other estimates.

3.2 Presence of difficult regions

When a gap is present in the grid, most smoothing methods just smooth across and over the gap. Smoothing across the gap is not allowed in the context of difficult regions, see Ramsay (2002). To deal with the presence of a gap in the domain, we consider the entries of the gap as *missing observations of type 2*. We reformulate the spline models such that we do not obtain estimates over the missing observations of type 2. The introduction of these pseudo-points also leads to the result that points from opposite sides of the gap do not depend on each other.

For smoothing on one-dimensional grids, let $D_2 = [\tau_1^*, \ldots, \tau_2^*]$ be the collection of indices corresponding to the points of the gap. When we are interested in the weighting patterns for the signal extraction of the spline at indices after the gap, i.e. $i > \tau_2^*$, we wish to have that both the filtered and smoothed spline do not depend on the gap and all values before the gap

$$w_t(a_{i|i-1}) = 0, \qquad w_t(a_{i|n}) = 0, \qquad t = 1, \ldots, \tau_2^*, \quad i > \tau_2^*. \tag{9}$$

Similarly, when we are computing the weighting patterns for smoothing at indices before the gap, i.e. $i < \tau_1^*$, zero-valued weights should be assigned to all points in the gap and to those observed after the gap

$$w_t(a_{i|n}) = 0, \qquad t = \tau_1^*, \ldots, n, \quad i < \tau_1^*. \tag{10}$$

No restrictions are imposed on the weights for filtering, since they are not affected by the presence of the gap. In Appendix B, we show that by setting the variance of the irregular term in the measurement equation to κ, and by re-initializing the elements in the state α_i in (4) at the points which correspond to the difficult region, i.e.

$$G_t = \kappa, \qquad \mathrm{Var}(\alpha_t) = \kappa I, \qquad t \in D_2, \tag{11}$$

we obtain weighting patterns which satisfy the conditions in (9) and (10). Furthermore, we show in Appendix C that by reformulating the model as given by (11), the variances of the filtered and smoothed estimates on D_2 are equal to κ

$$\mathrm{Var}(a_{t|t-1}) = \kappa I, \qquad \mathrm{Var}(a_{t|n}) = \kappa I, \qquad t \in D_2. \tag{12}$$

When the uncertainty on D_2 is large, we can discard estimates which are obtained for this set of points.

When we smooth on two-dimensional grids, we use missing observations of type 2 to represent the difficult region. No smoothing should take place over and across the difficult region. In each of the two steps of our approach, we reformulate the spline models as in the case of smoothing on a one-dimensional grid. In the first and second step, we assign zero-valued weights to all missing observations of type 2 and to all observational values located from the other side of the gap. As a result, to form our final estimate of the two-dimensional function, we do not use these pseudo points. Moreover, the pseudo points can also be used to create a rectangular grid. When the domain is non-rectangular, adding missing observations of type 2 leads to a rectangular domain.

4 Illustrations

4.1 A first illustration: motorcycle data

The motorcycle dataset consists of observations of acceleration against time (measured in milliseconds). The dataset is used for illustrative purposes by Silverman (1985) and is also employed by Koopman & Harvey (2003). The data is a cross section rather than a time series. The observations are not measured at an equidistance and at some points multiple values are recorded. Although it is not strictly necessary, they are removed from the dataset for practical reasons. We add missing observations of type 1 to create a regular grid, resulting in a new series which consists of 277 points, see panel (i) of Figure 3. All computations for this paper are done by using the object-oriented matrix language `Ox` of Doornik (2007).

The smoothing parameter $\lambda = \sigma_\epsilon^2/\sigma_\zeta^2$ is estimated by maximum likelihood and is given by $\hat{\lambda} = 0.102$. Smoothed estimates of the spline function are shown in panel (i) of Figure 3. We are interested in the weighting patterns, which are used to extract the smoothed spline at indices $i = 67$ and $i = 85$, which are shown in panels (ii) and (iii) of Figure 3, respectively. We see that larger values are assigned to observations close to the points of interest, while smaller weights are assigned to observations located further away. The gaps in the weighting patterns are caused by the presence of missing observations of type 1 since zero weights must be assigned to these pseudo points. These results are obtained implicitly by employing our methods.

Suppose we have a gap in the domain located in $D_2 = [69, \ldots, 78]$, see the shaded area in panel (i) of Figure 4. The gap is considered as a sequence of missing observations of type 2. We re-estimate the parameters of the spline model, resulting in an estimated smoothing parameter of 0.052. The smoothed estimates of the data are presented in panel (i) of Figure 4. In panel (ii) we depict the weighting pattern to form the smoothed state at $i = 67$. It

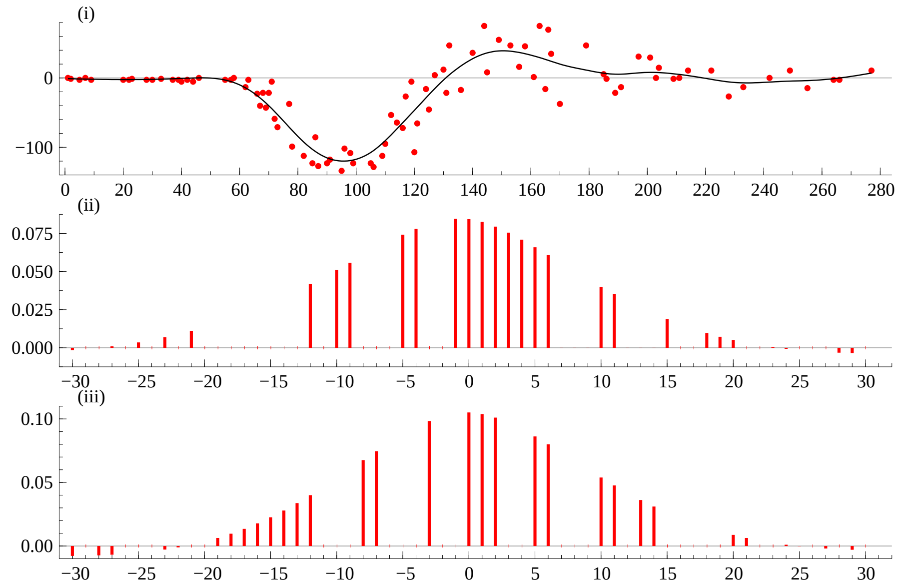

Figure 3: Smoothing on one-dimensional grid. The data and smoothed estimates of the smoothed spline are shown in panel (i). Panel (ii) presents the weighting pattern for extracting the smoothed spline at $i = 67$. In panel (iii) we plot the weighting pattern for extracting the smoothed spline at $i = 85$.

has zero weights for all values at indices $i \geq 69$. This implies that the gap itself and values observed after the gap are not used to form the state at $i = 67$. Moreover, the weighting pattern of the smoothed state at $i = 85$ has zero weights for all values at indices $i \leq 78$, see panel (iii).

By introducing two types of missing observations, we can handle irregular grids and assure that no smoothing takes place across and over the gap. Pseudo points are included to create a regular grid and they do not affect the estimates, since zero weights are assigned to these points. The results show that smoothed estimates neither depend on the gap nor on observations from the other side of the gap. Moreover, we do not have estimates for the gap as estimated values in the gap can be discarded due to the large variance.

4.2 The difficult region problem of Ramsay

Here we consider an illustration of smoothing data on two-dimensional grids. This illustration is motivated by the original difficult region problem of Ramsay (2002) and is also explored by Wang & Ranalli (2007). The data is simulated on a U-shaped grid and is represented

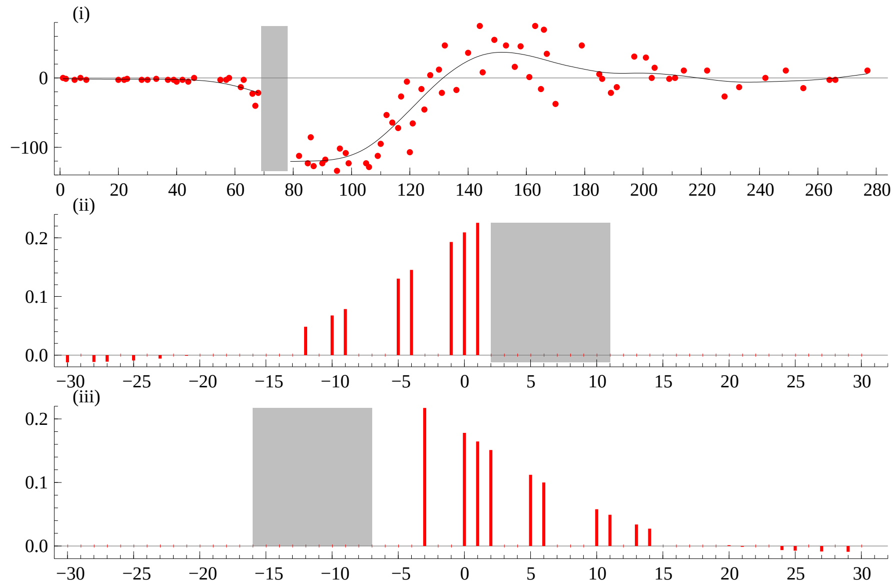

Figure 4: Motorcycle data: missing observations of type 2 are at $D_2 = [69, \ldots, 78]$, which is represented by the grey areas in all panels. The data and smoothed estimates of the smoothed spline are shown in panel (i). Panel (ii) presents the weighting pattern for extracting the smoothed spline at $i = 67$. In panel (iii) we plot the weighting pattern for extracting the smoothed spline at $i = 85$.

in Figure 5. The left-hand leg of the U-shaped function is sloping downwards, while the right-hand leg of the U-shape is sloping upwards. When we smooth the data, estimates on the left-hand leg should not depend on values observed from the right-hand leg and the other way around. Panel (ii) of Figure 5 views the U-shaped domain from the above. Since the function is not defined in the gap between the two legs, no smoothing should take place over the gap. The simulated data is obtained by adding noise, generated from a normal distribution $N(0, 1)$, to the true function, which is shown in panel (i) of Figure 6. To smooth the data, we apply the two-step smoothing method. The smoothing parameters of the model are estimated by maximum likelihood, i.e. on each slice of data we estimate the model parameters. The smoothed estimate of the U-shaped function from the simulated data is presented in panel (ii) of Figure 6. We observe that it closely resembles the true function. Furthermore, we observe that no estimates are obtained for all missing observations of type 2.

To show that our method does not smooth over and across the gap, we depict the weighting patterns corresponding to smoothing at two points located on the legs of the U-shape. We

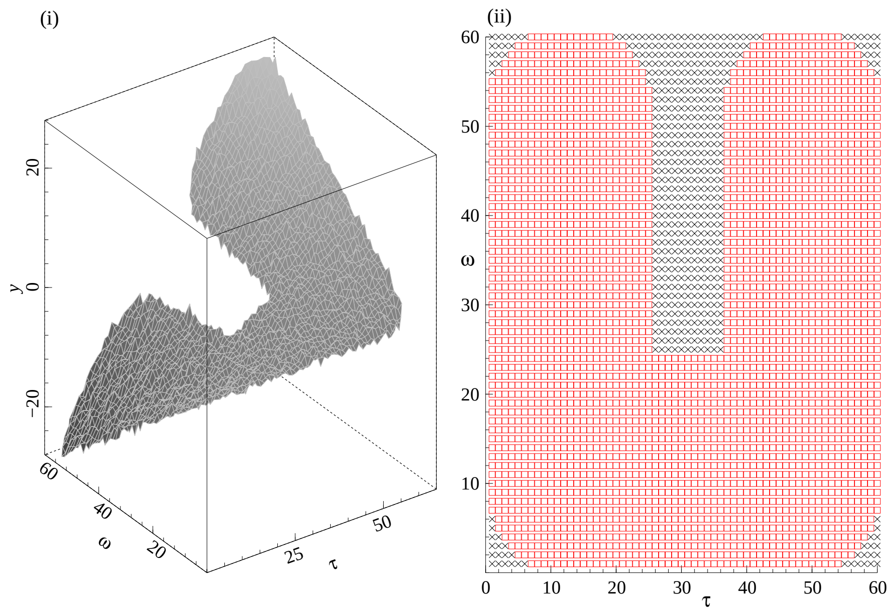

Figure 5: Panel (i) shows the data of the U-shaped domain example, which is also employed by Ramsay (2002). The left-hand leg slopes downward and the right-hand leg slopes upward. Panel (ii) views the U-shaped domain from above. Missing observations of type 2 are represented by the crosses. They correspond to the gap or to the points which are added to create a rectangular domain.

are interested in the weights to form the smoothed estimates at gridpoints (15, 30) and (50, 45), which lie on the left-hand and right-hand leg of the U-shape, respectively. The weighting patterns are graphically presented in Figure 7. Panel (i) corresponds to the point (15, 30) and panel (ii) is associated to the point (50, 45). The grey area in the figure represents the missing observations of type 2. In both panels, zero-valued weights are assigned to these pseudo-points, which implies that estimates do not depend on the gap or pseudo-points, which are added to form a rectangular grid. Additionally, to form estimates on one leg, we do not use observations from the other leg. The method only assigns non-zero weights to observational values located at the same leg as the point of interest.

4.3 A landscape problem

The second illustration of smoothing on two-dimensional grids is related to geographical problems such as a landscape with a lake or a mountain, which can disorder the composition of the landscape. The hole is regarded as a difficult region. Specific characteristics, e.g.

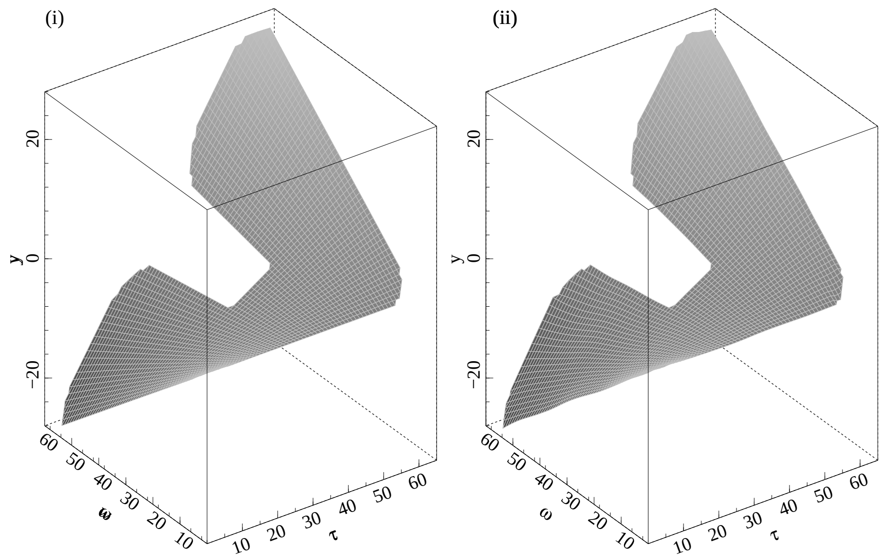

Figure 6: Panel (i) shows the true function $\mu(\tau, \omega)$ of the U-shape example on a 60×60 grid. Panel (ii) shows the smoothed estimates of the function $\mu(\tau, \omega)$ on the two-dimensional domain. Estimates are obtained by the two-step smoothing spline method. On each slice of the grid we estimate the model parameters.

income of a city located on one side of the lake may depend on cities located closely, but may not depend on a city which is located on the other side of the lake. Estimates on one side of the hole should therefore not depend on observations available from the other sides of the hole. This geographical problem is motivated by the example on income in the island of Montreal presented in Ramsay (2002). In that example there are holes in the domain which represent the airport and the water purification plant on the island. Over these areas, we do not wish to obtain estimates. Panel (i) of Figure 8 shows the simulated data on a 30×30 grid. Additionally, we have more difficulties by creating an irregular grid. In panel (ii) of Figure 8, we see the domain from above. The crosses in the figure represent the gap, while the black squares are associated to the irregularity of the grid.

The true function of the landscape example is shown in panel (i) of Figure 9. The data is simulated by adding noise, generated from a normal distribution $N(0, 1)$, to the true function. The model parameters are estimated on every slice of data in the two stages. In panel (ii) of Figure 9, we show the smoothed estimates. The proposed method smoothes around the difficult region and does not produce estimates in the hole. Moreover, we obtain estimates

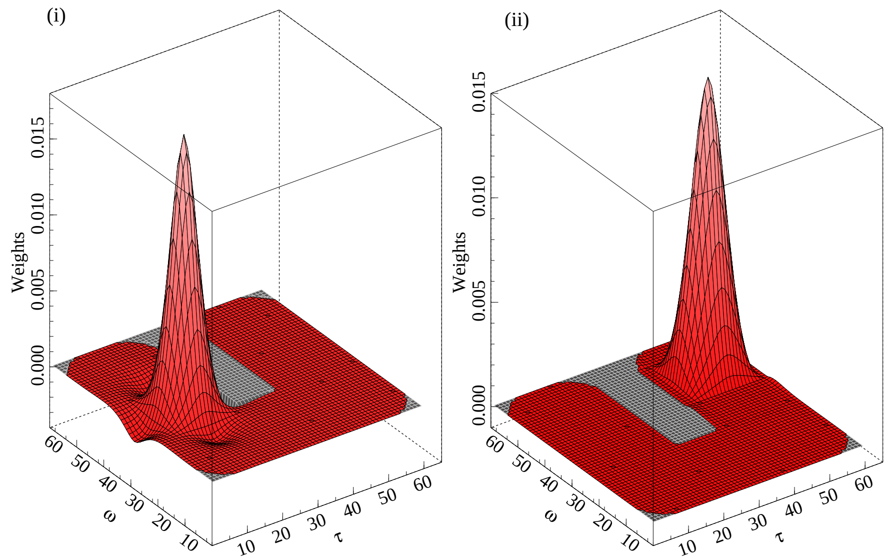

Figure 7: Weighting patterns associated to the smoothed function on the U-shaped domain. The grey area represents missing observations of type 2. Panels (i) shows the weighting pattern associated to the smoothed estimate at $(15, 30)$, which lies on the left-hand leg of the U-shape. In panel (ii) we depict the weights corresponding to smoothing at $(50, 45)$, which lies on the right-hand leg of the U-shape. The two-step smoothing spline method does not smooth across the gap, since estimates do not depend on observational values from the other side of the gap.

over the missing observations of type 1, which are added to create a regular grid. This proves that our approach can handle irregularly spaced observations. From the weighting patterns, we draw the same conclusions as in the case of the U-shape example. Values from the opposite side of the hole are not used to form smoothed estimates. We have zero-valued weights for missing observations of type 1 and type 2.

5 Conclusion

Smoothing data on one- or two-dimensional grids is not straightforward when irregularities are present in the grid. Many standard methods of smoothing cannot be used when the data is observed on irregular grids. Other problems arise when there is a gap or hole in the domain in which the data is not observed, since standard methods tend to smooth over and across the gap. Our solution to these problems is to adopt state space methods and introduce two types

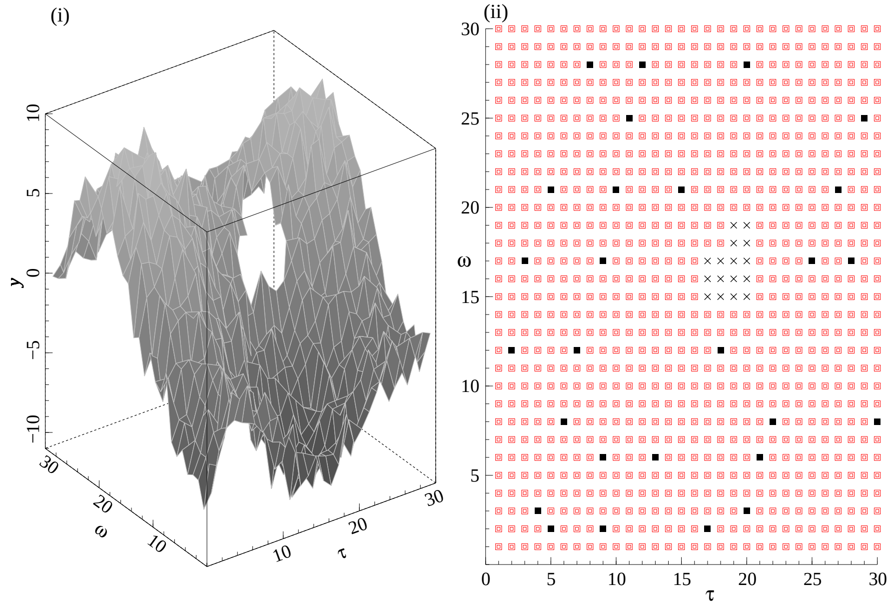

Figure 8: Panel (i) shows the data of the landscape example. Panel (ii) views the U-shaped domain from above. Black squares are associated to the irregularity of the grid and are considered as missing observations of type 1. Points corresponding to the hole are considered as missing observations of type 2 and are represented by the crosses.

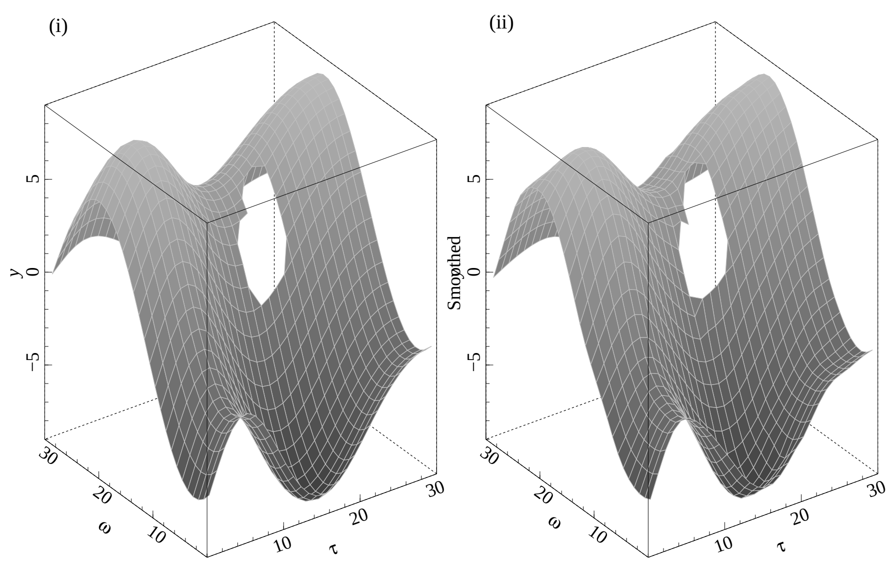

Figure 9: Panel (i) shows the true function $\mu(\tau, \omega)$ of the landscape example on a 60×60 grid. Panel (ii) shows the smoothed estimates of the function $\mu(\tau, \omega)$ on the two-dimensional domain. Estimates are obtained by the two-step smoothing spline method. On each slice of the grid we estimate the model parameters.

of missing observations. For the first type, we can handle irregularly spaced data. For the second type, we solve problems which are caused by the presence of gaps. By reformulating the spline models in state space, we can still apply standard smoothing methods when missing observations of type 1 or 2 are encountered.

To illustrate our methodology, we apply it to data observed on one- and two-dimensional grids. In the case of a one-dimensional grid, the data is not equally spaced and it contains a gap. We create a regular spaced grid by adding missing observations of type 1 and the gap is treated as a region of missing observations of type 2. It appears that our method does not smooth across or over the gap. When the data is observed on two-dimensional grids, we have the additional problem that the grid may not be rectangular. By adding missing observations of type 2 to the grid, we can tranform the grid into a rectangular one. We propose a two-step smoothing method to smooth the data on two-dimensional grids. Our method is tested on two examples. In both cases we show that our solution is effective in handling irregular and non-rectangular grids. Moreover, it does not produce estimates across and over difficult regions. The weighting patterns show zero-valued weights for points corresponding to the two types of missing observations and for observations from the other side of the gap.

We further like to emphasize that the necessary modifications for our computations are relatively small. Once the underlying smoothing model is reformulated in state space form, standard Kalman filtering and smoothing methods can be employed. In many software environments, KFS methods are available as a standard tool.

Appendix

A State space methods

We consider the following univariate linear state space model:

$$y_i = Z_i\alpha_i + \epsilon_i, \quad \epsilon_i \sim \text{NID}(0, \sigma_\epsilon^2), \quad i = 1, \ldots, n, \tag{A.1}$$

$$\alpha_{i+1} = T_i\alpha_i + \eta_i, \quad \eta_i \sim \text{NID}(0, H_i), \quad \alpha_1 \sim \text{NID}(a, P), \tag{A.2}$$

where y_i is the observation at index i, α_i is the unobserved state, ϵ_i and η_i are the disturbances in the measurement equation (A.1) and the transition equation (A.2) respectively. We assume that the disturbances are serially and mutually uncorrelated. The initial state vector is to have mean a and variance P. When $P = \kappa I$, $\kappa \to \infty$, we have a diffuse initialisation of the state vector. In practice we take $\kappa = 10^7$. The deterministic matrices T_i, Z_i and H_i are referred to as system matrices, which are in our case sparse selection matrices and contain the model parameters. Estimates of the mean and the variance of the state can be obtained by applying the Kalman filter and smoother, see Durbin & Koopman (2001). Denote the filtered state at i by $a_{i|i-1}$ and its variance by $P_{i|i-1}$. The Kalman filter equations are given by:

$$\begin{aligned} v_i &= y_i - Z_i a_{i|i-1}, & F_i &= Z_i P_{i|i-1} Z_i' + \sigma_\epsilon^2, & \\ K_i &= T_i P_{i|i-1} Z_i' F_i^{-1}, & L_i &= T_i - K_i Z_i, & i = 1, \ldots, n, \\ a_{i+1|i} &= T_i a_{i|i-1} + K_i v_i, & P_{i+1|i} &= T_i P_{i|i-1} L_i' + H_i, & \end{aligned} \tag{A.3}$$

where v_i is the one-step-ahead prediction error, F_i is its covariance matrix and K_i is known as the Kalman gain. The filtered state at index $i = 1$ is initialised by $a_{1|0} = a$ and $P_{1|0} = P$.

Smoothed estimates of the state $a_{i|n}$, are obtained by running the Kalman filter and subsequently the backwards recursion:

$$\begin{aligned} r_{i-1} &= Z_i' F_i^{-1} v_i + L_i' r_i, & N_{i-1} &= Z_i' F_i^{-1} Z_i + L_i' N_i L_i, & \\ a_{i|n} &= a_{i|i-1} + P_{i|i-1} r_{i-1}, & V_i &= P_{i|i-1} - P_{i|i-1} N_{i-1} P_{i|i-1}, & i = n, \ldots, 1, \end{aligned} \tag{A.4}$$

with the initialisation $r_n = 0$ and $N_n = 0$. The smoothed state estimate of α_i is denoted by $a_{i|n}$ with its mean square error matrix denoted by V_i.

Maximum likelihood estimates of the model parameters are obtained by maximising the Gaussian likelihood function. The Kalman recursions can be used to compute the likelihood function, which is given by

$$\log L(\theta) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{n}\left(\log|F_i| + v_i' F_i^{-1} v_i\right), \tag{A.5}$$

where the parameters are collected in θ and v_i, F_i are from the Kalman filter equations.

B Computing weights of the spline model

The state space form of the cubic spline model is presented in section 2.1. In this part of the appendix we derive the weights for filtering and smoothing. We consider two types of missing observations. The Kalman filter and smoother should obtain estimates over the missing observations of type 1, whereas over missing observations of type 2 we wish not to have estimates. Assume for simplicity that we have a single missing observation at τ_1^*. We can easily extend the case of a single missing observation to a sequence of missing observations. However, this does not change the derivation of the weights for filtering and smoothing. When we deal with a missing observation of type 1, we set

$$G_i = \begin{cases} \sigma_\epsilon^2 & \text{for } i \neq \tau_1^* \\ \kappa & \text{for } i = \tau_1^* \end{cases}, \tag{B.1}$$

such that

$$F_{\tau_1^*} = \kappa, \quad K_{\tau_1^*} = 0, \quad \kappa \to \infty.$$

In the case of the presence of a missing observation of type 2 at index τ_1^*, we set

$$G_{\tau_1^*} = \kappa, \quad P_{\tau_1^*+1} = \kappa I, \tag{B.2}$$

and we have

$$F_{\tau_1^*} = \kappa, \quad K_{\tau_1^*} = 0, \quad L_{\tau_1^*} = T,$$

where $\kappa \to \infty$.

B.1 Computing weights for filtering

Suppose that we are interested in computing the weights to form the filtered state at index i. We implicitly assign weights to the observations $y_t,\ t = 1, \ldots, i-1$:

$$a_{i|i-1} = \sum_{t=1}^{i-1} w_t(a_{i|i-1}) y_t, \tag{B.3}$$

where $w_t(a_{i|i-1})$ is the weight assigned to observation y_t. We assume further that a missing observation is found at index τ_1^*, where $\tau_1^* < i$. The case where $\tau_1^* > i$ is not interesting since the presence of the missing observation does not affect the filtering weights for computing $a_{i|i-1}$.

Recall from Koopman & Harvey (2003) that the weights for filtering are given by the following:

$$w_t(a_{i|i-1}) = B_{i,t} K_t, \quad t = i-1, \ldots, 1, \tag{B.4}$$

where

$$B_{i,i-1} = I, \quad B_{i,t} = L_{i-1} L_{i-2} \ldots L_{t+1}, \quad j = i-2, \ldots, 1. \tag{B.5}$$

Alternatively, $B_{i,t}$ can also be computed efficiently by the following backward recursion:

$$B_{i,t-1} = B_{i,t} L_t, \quad t = i-1, \ldots, 1. \tag{B.6}$$

Missing observation type 1

When we deal with a missing observation of type 1 at τ_1^*, the weight associated to the missing value is obviously zero:

$$w_{\tau_1^*}(a_{i|i-1}) = B_{i,\tau_1^*} K_{\tau_1^*} = 0, \tag{B.7}$$

since $K_{\tau_1^*} = 0$. Weights associated to other observations are computed by equation (B.4). We conclude that the filtered state does not depend on the missing value at τ_1^*.

Missing observation type 2

We adapt the model such that the Kalman filter does not filter across the missing observation, i.e. we need to show that the weights associated to the values at $t = 1, \dots, \tau_1^*$ are equal to zero. In order to show this we need the following at $\tau_1^* + 1$:

$$
\begin{aligned}
P_{\tau_1^*+1} &= \begin{pmatrix} \kappa & 0 \\ 0 & \kappa \end{pmatrix}, \\
F_{\tau_1^*+1} &= ZP_{\tau_1^*+1}Z' + \sigma_\epsilon^2 = \kappa, \\
K_{\tau_1^*+1} &= TP_{\tau_1^*+1}Z'F^{-1}_{\tau_1^*+1} = \begin{pmatrix} 1 & \delta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \kappa & 0 \\ 0 & \kappa \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \frac{1}{\kappa} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\
L_{\tau_1^*+1} &= T - K_{\tau_1^*+1}Z = \begin{pmatrix} 1 & \delta \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \delta \\ 0 & 1 \end{pmatrix},
\end{aligned}
\tag{B.8}
$$

where $\kappa \to \infty$. In the same way, we derive the matrices at $\tau_1^* + 2$:

$$
\begin{aligned}
P_{\tau_1^*+2} &= TP_{\tau_1^*+1}L'_{\tau_1^*+1} + H = \begin{pmatrix} 1 & \delta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \kappa & 0 \\ 0 & \kappa \end{pmatrix} \begin{pmatrix} 0 & 0 \\ \delta & 1 \end{pmatrix} + H \overset{\kappa\to\infty}{=} \begin{pmatrix} \delta^2\kappa & \delta\kappa \\ \delta^2\kappa & \kappa \end{pmatrix}, \\
F_{\tau_1^*+2} &= ZP_{\tau_1^*+2}Z' + \sigma_\epsilon^2 = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \delta^2\kappa & \delta\kappa \\ \delta^2\kappa & \kappa \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \sigma_\epsilon^2 \overset{\kappa\to\infty}{=} \delta^2\kappa, \\
K_{\tau_1^*+2} &= TP_{\tau_1^*+2}Z'F^{-1}_{\tau_1^*+2} = \begin{pmatrix} 1 & \delta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \delta^2\kappa & \delta\kappa \\ \delta^2\kappa & \kappa \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \frac{1}{\delta^2\kappa} = \begin{pmatrix} 2 \\ 1/\delta \end{pmatrix}, \\
L_{\tau_1^*+2} &= T - K_{\tau_1^*+2}Z = \begin{pmatrix} 1 & \delta \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 2 \\ 1/\delta \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & \delta \\ -1/\delta & 1 \end{pmatrix},
\end{aligned}
\tag{B.9}
$$

where H and σ_ϵ^2 vanish when $\kappa \to \infty$.

Using equations (B.5), (B.6) and the result that $L_{\tau_1^*+2}L_{\tau_1^*+1} = 0$, we have

$$B_{i,t} = 0, \quad t = 1, \dots, \tau_1^*,$$

and consequently

$$w_t(a_{i|i-1}) = 0, \quad t = 1, \dots, \tau_1^*, \tag{B.10}$$

We conclude that the weights for filtering associated to the observations at $t = 1, \dots, \tau_1^*$ are equal to zero. This implies that the Kalman filter does not filter across the missing observation of type 2, i.e. the filtered state at $i > \tau_1^*$ does not depend on values observed before the missing value.

B.2 Computing weights for smoothing

We derive in this section the weights for smoothing. To form the smoothed state $a_{i|n}$, we assign weights to all observations of the sample:

$$a_{i|n} = \sum_{t=1}^{n} w_t(a_{i|n}) y_t, \tag{B.11}$$

where $w_t(a_{i|n})$ is the weight assigned to observation y_t. The weights for smoothing are given by, see Koopman & Harvey (2003):

$$w_t(a_{i|n}) = \begin{cases} (I - P_{i|i-1} N_{i-1}) w_t(a_{i|i-1}), & \text{for } t = 1, \ldots, i-1, \\ B^*_{i,t} C_t, & \text{for } t = i, \ldots, n, \end{cases} \tag{B.12}$$

where $w_t(a_{i|i-1})$ is the weight for filtering from (B.4) and

$$\begin{aligned} C_t &= Z' F_t^{-1} - L_t' N_t K_t, \\ B^*_{i,t+1} &= B^*_{i,t} L_t', \end{aligned} \qquad t = i, \ldots, n, \tag{B.13}$$

with $B^*_{i,i} = P_{i|i-1}$. As in the case of the filtered weights we assume that the missing observation is found at index τ_1^*. We consider two cases of the location of the missing observation, i.e. $i > \tau_1^*$ and $i < \tau_1^*$.

B.2.1 Case where $i > \tau_1^*$

We first derive the weights for smoothing for the case $i > \tau_1^*$. This means that the presence of the missing observation only affects the weights at $t = 1, \ldots, \tau_1^*$. Other weights can be computed by equation (B.12).

Missing observation type 1

The presence of missing observation of type 1 only affects the weight at τ_1^*:

$$w_{\tau_1^*}(a_{i|n}) = (I - P_{i|i-1} N_{i-1}) w_{\tau_1^*}(a_{i|i-1}) = 0, \tag{B.14}$$

since $w_{\tau_1^*}(a_{i|i-1}) = 0$; see also the result in (B.7). Weights associated to other observations can be computed by equation (B.12).

Missing observation type 2

The weights for smoothing are given by

$$w_t(a_{i|n}) = \begin{cases} 0 & \text{for } t = 1, \ldots, \tau_1^*, \\ (I - P_{i|i-1}N_{i-1})w_t(a_{i|i-1}) & \text{for } t = \tau_1^* + 1, \ldots, i-1, \\ B_{i,t}^* C_t & \text{for } t = i, \ldots, n, \end{cases} \tag{B.15}$$

since $w_t(a_{i|i-1}) = 0$ for $t = 1, \ldots, \tau_1^*$; see the result in (B.7).

B.2.2 Case where $i < \tau_1^*$

The presence of the missing observation now only affects the weights associated to the missing observation and to the observations which appear after the missing observation, i.e. for $t = \tau_1^*, \ldots, n$.

Missing observation type 1

The missing observation of type 1 only affects the weight corresponding to i_1:

$$w_{\tau_1^*}(a_{i|n}) = B_{i,\tau_1^*}^* C_{\tau_1^*} = 0, \tag{B.16}$$

since

$$C_{\tau_1^*} = Z' F_{\tau_1^*}^{-1} - L_{\tau_1^*}' N_{\tau_1^*} K_{\tau_1^*} = 0,$$

where $K_{\tau_1^*} = 0$ and $F_{\tau_1^*}^{-1} = \kappa^{-1} = 0$ when $\kappa \to \infty$. Weights associated to other values can be computed by equation (B.12).

Missing observation type 2

When we encounter a missing observation op type 2 at $\tau_1^* > i$, the weights for smoothing are given by

$$w_t(a_{i|n}) = \begin{cases} (I - P_{i|i-1}N_{i-1})w_t(a_{i|i-1}) & \text{for } t = 1, \ldots, i-1, \\ B_{i,t}^* C_t & \text{for } t = i, \ldots, \tau_1^* - 1, \\ 0 & \text{for } t = \tau_1^*, \ldots, n. \end{cases} \tag{B.17}$$

To derive the weights we need the following intermediate results:

$$F_{\tau_1^*+1} = \kappa, \quad F_{\tau_1^*+2} = \delta^2\kappa, \quad L_{\tau_1^*+1} = \begin{pmatrix} 0 & \delta \\ 0 & 1 \end{pmatrix}, \quad L_{\tau_1^*+2} = \begin{pmatrix} -1 & \delta \\ -1/\delta & 1 \end{pmatrix}.$$

We derive the smoothing weights as follows:

- For $t = \tau_1^*$, we have

$$w_{\tau_1^*}(a_{i|n}) = B^*_{i,\tau_1^*} C_{\tau_1^*} = 0,$$

since $C_{\tau_1^*} = 0$

- For $t = \tau_1^* + 1$, we have

$$w_{\tau_1^*+1}(a_{i|n}) = B^*_{i,\tau_1^*+1} C_{\tau_1^*+1} = 0,$$

since

$$\begin{aligned} C_{\tau_1^*+1} &= Z' F^{-1}_{\tau_1^*+1} - L'_{\tau_1^*+1} N_{\tau_1^*+1} K_{\tau_1^*+1} \\ &= Z' F^{-1}_{\tau_1^*+1} - L'_{\tau_1^*+1} Z' F^{-1}_{\tau_1^*+2} Z K_{\tau_1^*+1} - L'_{\tau_1^*+1} L'_{\tau_1^*+2} N_{\tau_1^*+2} L_{\tau_1^*+2} K_{\tau_1^*+1} \\ &= 0, \quad \text{when } \kappa \to \infty \end{aligned}$$

where we use $N_{\tau_1^*+1} = Z' F^{-1}_{\tau_1^*+2} Z + L'_{\tau_1^*+2} N_{\tau_1^*+2} L_{\tau_1^*+2}$. Notice that

$$L'_{\tau_1^*+1} L'_{\tau_1^*+2} = 0, \quad Z' F^{-1}_{\tau_1^*+1} = 0, \quad L'_{\tau_1^*+1} Z' F^{-1}_{\tau_1^*+2} Z = 0 \quad \text{when } \kappa \to \infty.$$

- For $j = \tau_1^* + 2$, we have

$$\begin{aligned} w_{\tau_1^*+2}(a_{i|n}) &= B^*_{i,\tau_1^*+2} C_{\tau_1^*+2} \\ &= B^*_{i,\tau_1^*+1} L'_{\tau_1^*+1} C_{\tau_1^*+2} \\ &= B^*_{i,\tau_1^*+1} L'_{\tau_1^*+1} (Z' F^{-1}_{\tau_1^*+2} - L'_{\tau_1^*+2} N_{\tau_1^*+2} K_{\tau_1^*+2}) \\ &= 0, \quad \text{when } \kappa \to \infty \end{aligned}$$

since $L'_{\tau_1^*+1} Z' F^{-1}_{\tau_1^*+2} = 0$, $\kappa \to \infty$ and $L'_{\tau_1^*+1} L'_{\tau_1^*+2} = 0$.

- For $t = \tau_1^* + 3, \ldots, n$, it is straightforward to show that the weights are equal to zero

$$\begin{aligned} w_t(a_{i|n}) &= B^*_{i,t} C_t \\ &= B^*_{i,\tau_1^*+1} L'_{i,\tau_1^*+1} L'_{i,\tau_1^*+2} \cdots L'_{t-1} C_t \\ &= 0, \end{aligned}$$

since $L'_{i,\tau_1^*+1} L'_{i,\tau_1^*+2} = 0$ and we repeatedly use $B^*_{i,t+1} = B^*_{i,t} L'_{i,t}$.

C Variance of the smoothed state in the difficult region

Suppose we have a set of sequential missing observations of type 2. Collect the indices corresponding to these points in $D_2 = [\tau_1^*, \ldots \tau_2^*]$. In the derivations below we use the results in (B.8) and (B.9). We have

$$\begin{aligned} N_i &= Z'F_{i+1}^{-1}Z + L_{i+1}'N_{i+1}L_{i+1} \\ &= Z'F_{i+1}^{-1}Z + L_{i+1}'(Z'F_{i+2}^{-1}Z + L_{i+2}'N_{i+2}L_{i+2})L_{i+2} \\ &= Z'F_{i+1}^{-1}Z + L_{i+1}'Z'F_{i+2}^{-1}ZL_{i+2} + L_{i+1}'L_{i+2}'N_{i+2}L_{i+2}L_{i+2} \\ &= 0, \quad \text{when } \kappa \to \infty, \quad \text{for } i \in D_2, \end{aligned} \tag{C.1}$$

since $F_{i+1}^{-1} = 0$, $L_{i+1}'Z'F_{i+2}^{-1} = 0$ and $L_{i+1}'L_{i+2}' = 0$ for $i \in D_2$ and when $\kappa \to \infty$. The variance of the smoothed state $a_{i|n}$ is now given by:

$$\begin{aligned} V_i &= P_{i|i-1} - P_{i|i-1}N_{i-1}P_{i|i-1} \\ &= P_{i|i-1}, \qquad\qquad i = \tau_1^* + 1, \ldots \tau_2^*. \end{aligned} \tag{C.2}$$

Notice that this does not apply for $V_{\tau_1^*}$, since $N_{\tau_1^*-1} \neq 0$. Summarising, the diagonal elements of the variance matrix of the state vector at D_2, except for the first point of D_2, is equal to κ which goes to infinity. It is already obvious that the diagonal elements of the variance matrix of the filtered state is proportional to κ.

References

Ansley, C. F., Kohn, R. & Tharm, D. (1991), 'The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters', *Journal of the American Statistical Association* **86**, 1042–1050.

Ansley, C. F., Kohn, R. & Wong, D. (1992), 'Nonparametric spline regression with autoregressive moving-average errors', *Biometrika* **79**, 335–346.

Cressie, N. (1993), *Statistics for spatial data*, Wiley, New York.

de Jong, P. (1988), 'A cross-validation filter for time series models', *Biometrika* **75**, 594–600.

Doornik, J. A. (2007), *Ox: An Object-Oriented Matrix Language*, 1st edn, Timberlake Consultants Ltd, London.

Durbin, J. & Koopman, S. J. (2001), *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford.

Gao, F., Wahba, G., Klein, R. & Klein, B. (2001), 'Smoothing spline ANOVA for multivariate Bernoulli observations with application to ophthalmology data', *Journal of the American Statistical Association* **96**(453), 127–160.

Green, P. J. & Silverman, B. W. (1994), *Nonparametric regression and generalized linear models: a roughness penalty approach*, Chapman and Hall, London.

Gu, C. (2002), *Smoothing Spline ANOVA Models*, Springer Verlag, New York.

Gu, C. & Wahba, G. (1993), 'Semi-parametric analysis of variance with tensor product thin plate splines', *Journal of the Royal Statistical Society, Series B* **55**, 353–368.

Guo, W. (2002), 'Inference in Smoothing Spline ANOVA', *Journal of the Royal Statistical Society B* **64**, 887–898.

Guo, W., Dai, M., Omboa, H. C. & von Sachs, R. (2003), 'Smoothing Spline ANOVA for time-dependent spectral analysis', *Journal of the American Statistical Association* **98**, 643–652.

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Hastie, T. & Tibshirani, R. (1990), *Generalized Additive Models*, Chapman & Hall, London.

Horgan, G. W. (1999), 'Using wavelets for data smoothing: a simulation study', *Journal of Applied Statistics* **26**, 923–932.

Kammann, E. E. & Wand, M. P. (2003), 'Geoadditive models', *Journal of the Royal Statistical Society C* **52**, 1–18.

Kohn, R. & Ansley, C. F. (1987), 'A new algorithm for spline smoothing based on smoothing a stochastic process', *SIAM Journal of Science and Statistical Computing* **8**, 33–48.

Koopman, S. J. & Harvey, A. (2003), 'Computing observation weights for signal extraction and filtering', *Journal Economic Dynamics and Control* **27**, 1317–1333.

Qin, L. & Guo, W. (2006), 'State space representation for Smoothing Spline ANOVA models', *Journal of Computational and Graphical Statistics* **15**(4), 830–8473.

Ramsay, T. (2002), 'Spline smoothing over difficult regions', *Journal of the Royal Statistical Society B* **64**(2), 307–319.

Schweppe, F. (1965), 'Evaluation of likelihood functions for Gaussian signals', *IEEE Transactions on Information Theory* **11**, 61–70.

Silverman, B. W. (1985), 'Some aspects of the spline smoothing approach to non-parametric regression curve fitting', *Journal of the Royal Statistical Society B* **47**(1), 1–52.

Wahba, G. (1990), *Spline Models for Observational Data*, SIAM, Philadelphia.

Wand, M. P. & Jones, M. C. (1995), *Kernel Smoothing*, Chapman and Hall, London.

Wang, H. & Ranalli, M. G. (2007), 'Low-rank smoothing splines on complicated domains', *Biometrics* **63**(1), 209–217.

Wang, Y., Ke, C. & Brown, M. B. (2003), 'Shape-invariant modeling of circadian rhythms with random effects and Smoothing Spline ANOVA decompositions', *Biometrics* **59**(4), 804–812.

Wecker, W. E. & Ansley, C. F. (1983), 'The signal extraction approach to nonlinear regression and spline smoothing', *Journal of the American Statistical Association* **78**, 81–89.